

「モデルに基づいた地球統計学(model-based geostatistics)」による公示地価の補間：山梨県内の公示地価を例として

著者名(日)	倉澤 一孝
雑誌名	山梨学院大学現代ビジネス研究
巻	3
ページ	33-40
発行年	2010-02-01
URL	http://id.nii.ac.jp/1188/00000032/

「モデルに基づいた地球統計学(model-based geostatistics)」 による公示地価の補間

山梨県内の公示地価を例として

Kriging Posted Land Prices: A Model-Based Geostatistical Approach

倉 澤 一 孝

Kazutaka KURASAWA

【概要】

公示地価は土地の取引価格の指標となることを目的として公表されているが、公示対象となる標準地の数は少ない。本稿では、標準地以外の土地にも価格情報を提供する方法として「モデルに基づいた地球統計学 (model-based geostatistics)」を検討する。山梨県内の公示地価を例として取り上げ、県内地価の連続的な分布を統計学的に予測 (クリギング) する。なお、推定と予測には、パラメータの不確実性を予測に反映させるため、ベイズ法を用いる。

【キーワード】

地球統計学、クリギング、公示地価、不動産鑑定評価、ベイズ推定

1. はじめに

公示地価は土地の取引価格の指標となることを目的として公表されている^[1]。しかし、公示対象となる標準地の数は少なく、実際の土地取引で公示地価が利用できる場面は限られている。2009年に公示された山梨県内の公示地価を例にとると、標準地に選ばれたのは179地点のみである(図1)。土地取引を行う者にとって価格情報は重要であり、標準地以外の土地についても価格情報が提供されることが望ましい。また、不動産鑑定士が行う鑑定評価では、鑑定評価の対象となっている土地の試算価格と公示地価の整合性を検証することが求められるが、当該土地が標準地であることは稀である。現状では、近隣・類似地域の公示地価を不動産鑑定士が事情補正し、補正後の公示地価を代用して整

合性の検証が行われている。もし公示地価に準ずる価格情報が全ての土地で提供されれば、より精度の高い鑑定評価が期待できるであろう。

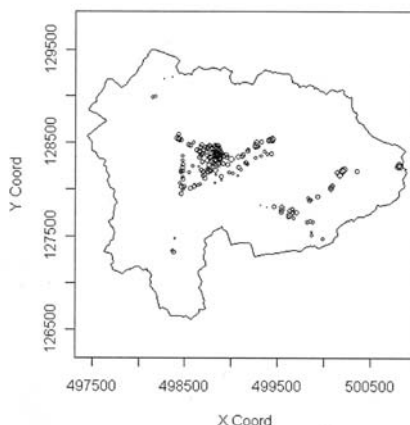


図1 山梨県内の標準地と公示地価。X Coordは経度、Y Coordは緯度を表す。なお、標準地を示す円の半径は公示地価に比例している。

では、標準地以外の土地で価格情報を提供するにはどうすればよいだろうか。解決策の一つとして統計学的に地価を補間(interpolation)することが考えられる。土地は一筆の土地が独立して存在するのではなく、近隣にある複数の土地が地域を構成し、近隣地域内の土地が相互に依存・競争・代替の関係を持ちながら存在している。土地の利用状況を観察してみると、空間的に近接する土地は類似した用途として利用されており、住宅地域、商業地域、工業地域などの用途的地域を構成している。土地にはこのような特性があるため、地価の空間的相関関係は強い。したがって、近隣地域の公示地価から標準地以外の地価を補間すれば、精度の高い予測値が得られると予想される。

公示地価などの空間データを補間する方法はいくつかあるが、有効な方法の一つが「モデルに基づいた地球統計学(model-based geostatistics)」である。この方法では、関心下の変数について確率モデルを作り、サンプル地点のデータからモデルのパラメータを推定する。そして、モデルから計算される予測値で欠損値を補間し、変数の連続的な分布を表現する。この方法は様々な分野で使われており、応用例には疾病地図の作成や有害物質の分布調査などがある[2]。

地球統計学(geostatistics)は空間上に点として表示されるデータ(point-referenced data)を解析するための手法であり、その予測法は「クリギング(kriging)」と呼ばれている。クリギングは、元々、Krige(1951)やMatheron(1963, 1971)などによって鉱床の空間分布を予測するために開発されたが、他の空間統計学の影響を受けずに独自に発展したため、数理統計学的基礎を欠いていた。後に、Watson(1971, 1972)やRipley(1981)などが数理統計学的観点からクリギングを再検討し、その結果、地球統計学は空間統計学の一分野とし

て位置づけられるようになった(Cressie(1993))。Diggle, Tawn and Moyeed(1998)は、数理統計学に基礎付けされた地球統計学を「モデルに基づいた地球統計学」と呼び、伝統的な地球統計学と区別することを提唱している。

本稿では、山梨県内の公示地価を例として取り上げ、「モデルに基づいた地球統計学」を応用して公示地価を補間する。以下、分析は次のように進める。次節では、県内地価の確率モデルを構築する。簡単化のために、地価は対数正規分布に従い、地価の相関係数は距離の関数であると仮定して、地価の分布をモデル化する。第3節では、モデルのパラメータの推定方法と地価の予測方法について解説する。地球統計学モデルを推定する方法はいくつかあるが、パラメータの不確実性を予測に反映させるため、ベイズ法を採用する。第4節では、2009年の公示地価からパラメータを推定し、標準地以外の土地の価格を予測する。モンテカルロ・シミュレーションでパラメータの事後分布と地価の事後予測分布を近似し、事後予測分布の平均値で欠損値を補間する。また、モデルの予測精度についても検証を行う。最終節では、本稿で行った分析についてまとめる。

2. 県内地価の確率モデル

県内地価の分布をモデル化したものとして、次の確率モデルを考える。

$$\log(Y_i) = S_i + N_i$$

このモデルでは、地点*i*における地価の対数 $\log(Y_i)$ はシグナル S_i とノイズ N_i に分解される[3]。シグナル S_i は「真の地価」を表し、正規分布 $N(\beta_i, \sigma^2)$ に従う。ノイズ N_i は測定誤差を表し、正規分布 $N(0, \tau^2)$ に従う。シグナル S_i とノイズ N_i が独立であると仮定すると、地価 Y_i は対数正規分布に従い、

$$\log(Y_i) \sim N(\beta_i, \sigma^2 + \tau^2)$$

となる。また、ベクトルは多変量正規分布に従い、

$$\log(Y) \sim MVN(\beta, \sigma^2 P + \tau^2 I)$$

となる。ただし、 β は平均のベクトル、 P は相関行列、 I は単位行列である。

平均のベクトル β は定数とするか、データに空間的なトレンドが見られる場合、一次関数や二次関数などを仮定する。空間的トレンドを探索的に特定するために、経度・緯度を横軸として地価の対数を図2にプロットした。県内の公示地価に空間的なトレンドは見られないので、平均のベクトル β は定数とするのが適当であると考えられる。

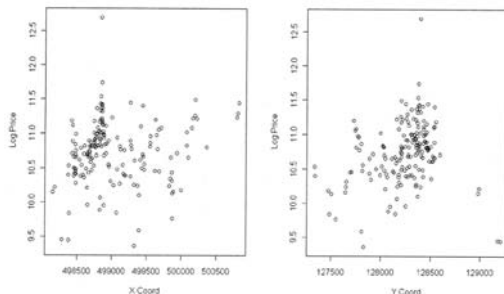


図2 経度(Y Coord)・緯度(X Coord)を横軸として地価の対数をプロットした。

相関行列 P の要素については、相関係数を距離の関数と仮定するのが一般的である。この仮定の下では、二地点の位置関係がどの方向であっても、距離が同じであれば相関係数は変わらない。これは「等方性(isotropy)」と呼ばれる。本稿でも等方性を仮定し、地点 i と地点 j の相関係数 ρ_{ij} は距離 d_{ij} の関数とする。相関関数には指数関数 $\rho_{ij} = \exp(-d_{ij}/\phi)$ を使う($\phi > 0$)[4]。なお、距離 d_{ij} は世界測地系の秒を単位として測る(緯度35度上で、経度1秒の長さは約25.0 m、緯度1秒の長さは約30.8 m)。

3. ベイズ推定と予測

地球統計学モデルのパラメータを推定する方法はいくつかあるが、本稿では、パラメータの不確実性を予測に反映させるため、ベイズ法を用いる(ベイズ推定・予測の方法を図3に示した)。ベイズ法では、パラメータを確率変数と考え、パラメータの確率分布をデータから推定する。推定された確率分布は「事後分布」と呼ばれ、パラメータの不確実性は事後分布の散布度に反映される。また、関心下の変数は「事後予測分布」として予測される。事後予測分布はパラメータの事後確率で予測値を加重して計算されるので、事後予測分布にもパラメータの不確実性が反映される。

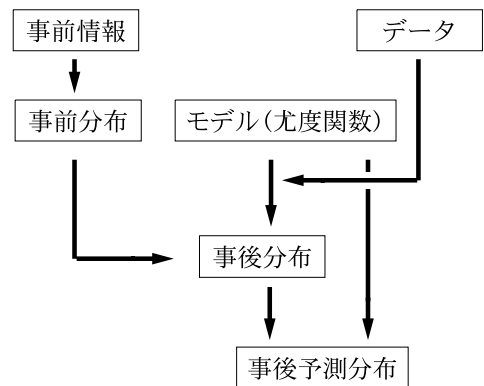


図3 ベイズ推定・予測の方法。

ベイズ法でパラメータを推定するときは、分析者が持っている事前情報を確率分布として表現する。この確率分布は「事前分布」と呼ばれる。調査・研究等の公的分析では事前情報をできるだけ排して分析するのが望ましいとされ、「漠然とした」確率分布が事前分布に使われることが多い[5]。本稿でも

$$p(\beta) = U(-\infty, \infty)$$

$$p(\sigma^2) \propto 1/\sigma^2$$

$$p(\tau^2/\sigma^2) = U(0,1)$$

$$p(\phi) = U(0, \infty)$$

を使用する。平均と β 指数関数の係数の ϕ 事前分布は一様分布、分散 σ^2 の事前分布は逆数 $1/\sigma^2$ に比例するものとする。また、測定誤差 τ^2 については、相対的な測定誤差 τ^2/σ^2 を定義し、この事前分布を一様分布とする。なお、各パラメータが独立であれば

$$p(\theta) = p(\beta)p(\sigma^2)p(\tau^2/\sigma^2)p(\phi)$$

となる。ただし、 $\theta \equiv (\beta, \sigma^2, \tau^2/\sigma^2, \phi)$ である。

次に、事前分布と尤度関数 (モデル) からベイズの法則によって事後分布を求める。地価モデルの尤度関数は

$$p(\log(Y) | \theta)$$

$$\propto |\Sigma| \exp\left(-\frac{1}{2}(\log(Y) - \beta)^t \Sigma^{-1}(\log(Y) - \beta)\right)$$

であるので、事後分布は

$$p(\theta | \log(Y)) \propto p(\log(Y) | \theta)p(\theta)$$

となる。ただし、 $\Sigma \equiv \sigma^2 P + \tau^2 I$ である。なお、事後分布の解析的な表現を求めるのが難しい場合、モンテカルロ・シミュレーションによって事後分布を近似し、平均、標準偏差、分位数等を数値計算する。

最後に、事後分布とモデルから、関心下の変数の事後予測分布を求める。地価の事後予測分布は

$$p(\log(\hat{Y}) | \log(Y)) \propto p(\log(\hat{Y}) | \theta)p(\theta | \log(Y))$$

となる。ただし、 $p(\log(\hat{Y}) | \theta)$ はモデル (尤度関数)、 \hat{Y} は地価の予測値である。事後分布と同様に、事後予測分布もモンテカルロ・シミュレーションで近似することができる。

4. パラメータの推定結果と公示地価の補間

以上の推定方法に従い、2009 年の公示地価から事後分布を推定した結果が表 1 と図 4 である。メトロポリス・ヘイスティング法でモンテカルロ・シミュレーションを行い、パラメータの集合についてサンプリングを 3000 回繰り返した [6] [7] [8]。各パラメータの平均値、標準偏差および分位数を表 1、ヒストグラムと近似曲線を図 4 に示した。

指数関数のパラメータ ϕ の推定結果を見ると、事後分布の平均値が 1719.44、中央値が

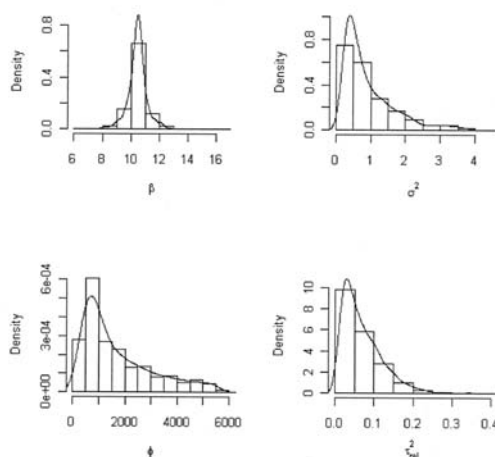


図 4 事後分布のヒストグラムと近似曲線。サンプリングした 3000 の値からヒストグラムを作成した。なお、 τ_{rel}^2 は相対誤差のことである。

	平均	標準偏差	分位数				
			2.5%	25.0%	50.0%	75.0%	97.5%
β	10.46	0.73	8.89	10.14	10.47	10.781.2	12.05
σ^2	0.93	0.76	0.19	0.38	0.65	6	3.14
τ^2/σ^2	0.07	0.05	0.01	0.03	0.06	0.10	0.21
ϕ	1719.44	1350.70	332.24	664.48	1218.21	2436.41	5205.07

表 1 事後分布の平均、標準偏差および分位数。

1218.21、95 パーセント確信区間が 332.24～5205.07 となっている。図 5 は、二地点間の距離が変わるにつれて相関係数がどのように変化するか見るために、平均値 1719.44 をパラメータ値とする指数関数をプロットしたものである。距離が長くなるとともに相関係数は小さくなり、二地点間の距離が 100 秒 (約 2.5～3.1 km) のときは 0.95 程度であるが、500 秒 (約 12.5～15.4 km) では 0.75 より小さくなり、

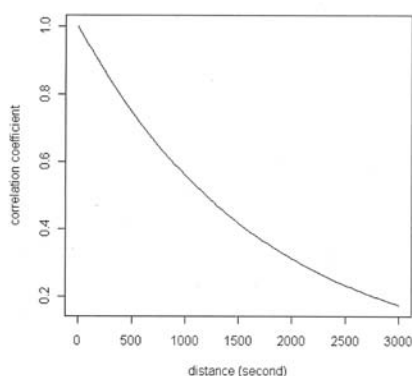


図 5 事後分布の平均値 1719.44 を ϕ のパラメータ値とする指数関数 $p_{ij} = \exp(-d_{ij}/\phi)$ をプロットした。

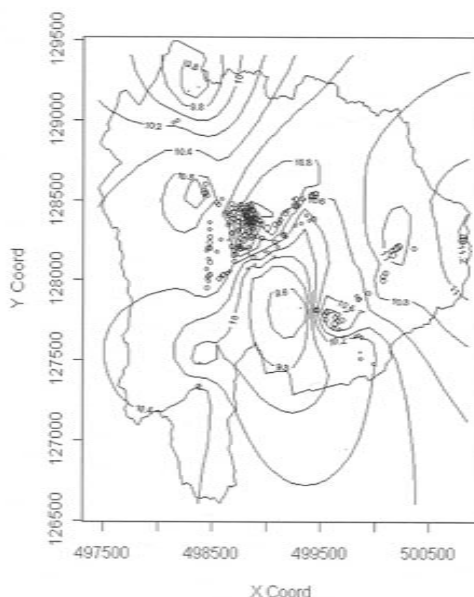
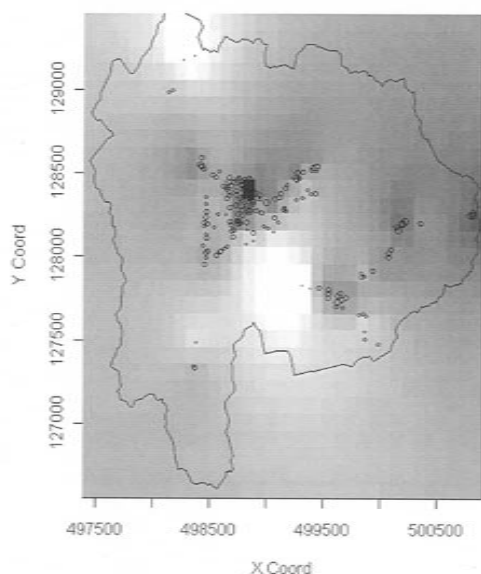


図 6 事後予測分布の平均値で公示地価を補間し、県内地価の分布を等高線グラフで表した。左図では、地価を等間隔で 10 段階に分け、グレー・スケールで等高線グラフを作成した。右図では、地価の値を等高線で示した。なお、図中の円は図 1 の標準地と同じものである。

1000 秒 (約 25.0～31.8 km) を超えると 0.60 を下回る。

また、相対的な測定誤差を表す τ^2/σ^2 は、事後分布の平均値が 0.07、中央値が 0.06、95 パーセント確信区間が 0.01～0.21 となっている。この推定結果から、測定誤差 τ^2 の大きさは、シグナルの分散 σ^2 に対して 1% から 21% 程度、地価の分散 $\sigma^2 + \tau^2$ に対して 1% から 17% 程度であると推測される。これは、鑑定誤差は無視できるほど小さいと言えず、「真の地価」が正確に測定されていない可能性を示唆している。

これらの推定結果をもとに、地価の事後予測分布から公示地価を補間した結果が図 6 である。補間には、事後予測分布の代表値として平均値を使用した。事後予測分布の数値計算では、経線・緯線に沿って 100 秒 (約 2.5～3.1 km) 間隔で格子を作成し、格子点ごとにモン

テカルロ・シミュレーションを行った。事後分布と同様に、メトロポリス・ヘイスティング法

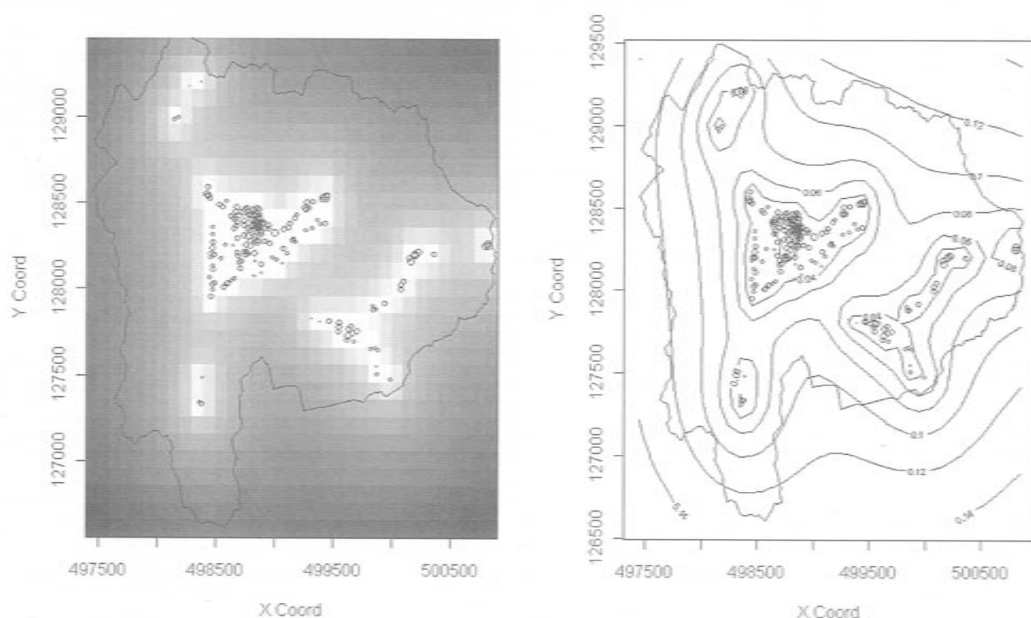


図7 事後予測分布の変動係数を等高線グラフで表した。左図では、変動係数を等間隔で10段階に分け、グレー・スケールで等高線グラフを作成した。右図では、変動係数の値を等高線で示した。なお、図中の円は図1の標準値と同じものである。

を使用し、各格子点で3000回ずつサンプリングを行った。

また、予測に対する確信度を評価するために、事後予測分布の変動係数を図7の等高線グラフに示した。変動係数は標準地（サンプル地点）から離れると大きくなる傾向が見られるが、最大でも0.12程度であり、予測に対する確信度は十分高いと言える。

最後に、この確率モデルの予測精度について検証を行った。179の標準地からランダムに143地点（データの約8割）を抽出してパラメータ推定を行い、残りの36地点（データの約2割）の地価を予測した^[9]。これを12回繰り返し、各回のパーセンテージ誤差をヒストグラムとして図8に示した。ほとんどの誤差は±5パーセント以下であり、±10パーセントを超える誤差は稀である。この検証結果から、モデルの予測精度は十分に高く、誤差5パー

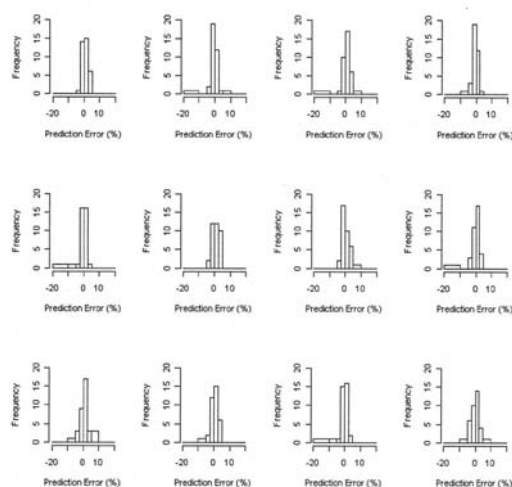


図8 予測精度を検証するために、各試行のパーセント予測誤差をヒストグラムに示した。

セント以下の精度で地価予測が可能であることが分かる。

5. むすびに

本稿では、対数正規性および等方性を仮定して地価の確率モデルを構築し、ベイズ法でパラメータの推定と地価の予測を行った。そして、事後予測分布の平均値で欠損値を補間し、公示地価の連続的分布を等高線グラフで表した。また、モデルの予測精度を検証した結果、予測誤差は十分に小さくなることが期待でき、「モデルに基づいた地球統計学」の手法は公示地価の補間に適用可能であることが確認された。

更に詳しい分析を行うためには、いくつかの点についてモデルを拡張することが考えられる。まず、本稿では平均を定数と仮定したが、これをヘドニック関数としてモデル化する方法もある。たとえば、市町村や用途地域のダミー変数や主要駅からの距離などを共変量としてモデル化すれば、地価の変動に関してより詳しい理論的考察を行うことができる。また、異時点のデータを利用した時空間(spatio-temporal)モデルを応用することも考えられる。データには複数年の公示地価や、都道府県が公表する基準地価などが利用できるであろう。このようなモデルの拡張が予測精度にどのような影響を与えるかについては、次の機会に検証することしたい。

注

- [1]地価公示法では、公示地価の役割として①一般の取引に対する指標の提供、②不動産鑑定士の鑑定評価の規準、③公共用地の取得価格の算定の規準が与えられている。
- [2]「モデルに基づいた地球統計学」の応用例については Diggle and Ribeiro Jr.(2007)を参照。また、ベイズ統計学の立場から地球統計学を応用した事例については Congdon(2001, 2003)が詳し

い。

- [3]地価は非負なので対数変換するのが一般的である。また、関心下の変数の分布によっては、ボックス・コックス変換した方がよい場合もある。このように、変換後の変数が正規分布に従うと仮定したモデルは「trans-Gaussian モデル」と呼ばれる。
- [4]ガウス型空間過程(Gaussian spatial process)において、分散が非負になるためには、共分散行列が正定値でなければならない、共分散関数は正定値関数でなければならない。指数関数はこの条件を満たす関数の一つである。
- [5]公的分析を行う場合の事前分布の選択については 渡部(1999)を参照。
- [6]メトロポリス・ヘイスティング法はマルコフ連鎖モンテカルロ・シミュレーションのアルゴリズムの一つである。詳しくは Gelman et al.(2004)を参照。
- [7]初期値の影響を取り除くため、定常状態に達するまでの値は除外した。
- [8]モンテカルロ・シミュレーションには統計ソフト R を使用した。R を使ったモンテカルロ・シミュレーションについては Albert(2007)を参照。
- [9]事後予測分布を使ったモデル検証の方法については Gelman et al.(2004)を参照。

参考文献

- [1] Albert, J.(2007). Bayesian Computation with R. Springer
- [2] Congdon, P. (2001). Bayesian Statistical Modelling. Wiley
- [3] Congdon, P. (2003). Applied Bayesian Modelling. Wiley
- [4] Cressie, N. (1993). Statistics for Spatial Data. Wiley
- [5] Diggle, P. J. and Ribeiro Jr, P. J. (2007). Model-based Geostatistics. Springer
- [6] Diggle, P. J., Tawan, J. A., and Moyeed, R. A. (1998). "Model-based Geostatistics". Applied Statistics 47, 299-350
- [7] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). Bayesian Data Analysis.

Chapman & Hall/CRB

- [8] Krige, D. G. (1951). “A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa 52, 119–139
- [9] Matheron, G. (1963). “Principles of Geostatistics”. Economic Geology 58, 1246–1266
- [10] Matheron, G. (1971). “The Theory of Regionalized Variables and its Applications”. Technical Report 5, Cahiers du Centre de Morphologie Mathematique.
- [11] Ripley, B. D. (1981). Spatial Statistics. Wiley
- [12] Watson, G. S. (1971). “Trend–Surface Analysis”. Mathematical Geology 3, 215–226
- [13] Watson, G. S. (1972). “Trend–Surface Analysis and Spatial Correlation”. Geology Society of America Special Paper 146, 39–46
- [14] 渡部洋(1999). 「ベイズ統計学入門」 福村出版
- [15] 森島義博(2005). 「不動産鑑定入門」 東洋経済新報社